

NiceTag Ontology: tags as named graphs

Freddy Limpens¹, Alexandre Monnin^{2,3}, David Laniado⁴, Fabien Gandon¹

¹ Edelweiss, INRIA Sophia-Antipolis

`{freddy.limpens, fabien.gandon}@sophia.inria.fr`

² EXeCO, Paris I Pantheon - Sorbonne University

`Alexandre.Monnin@malix.univ-paris1.fr`

³ DICEN, Conservatoire National des Arts et Metiers (CNAM)

⁴ DEI, Politecnico di Milano

`david.laniado@elet.polimi.it`

Abstract. Current tag modelling does not fully take into account the rich and diverse nature tags, as signs, can take on. We propose an ontology of tags in which tags are modelled as named graphs. These named graphs are made of a resource linked to a “sign” which can be any resource reachable on the Web (an ontology concept, an image, etc.). The purpose of our model is to be able to describe tags in a very general manner, and as an immediate consequence, to describe tags as modelled by other tag models (SCOT, CommonTag, etc.). Our tag model can thus be seen as a bridge between the manifold conceptualizations and instantiations of tags’s models currently available on the Web.

1 Introduction

Tags are nowadays a key feature of the social Web and a new form of expression that can serve many purposes: categorizing or classifying content, comment, vote, react, express, share, identify, etc. Social tagging and the resulting folksonomies can be seen as a new opportunity to involve users in a novel relationship with web content.

The goal of our model, drafted at the VoCamp in Nice 2009⁵, is to allow for modelling tag actions without being bound to a unique model either of the sign used to tag or of its semantics. In order to describe tags in the most flexible manner, we propose considering them primarily as a link between a tagged resource and a sign used to tag, which can take on many different forms and conceptualizations (an image, a literal, an ontology concept, etc.). These two entities are modelled with the `rdfs:Resource` class from the RDF specifications in order to leave the choice of the model of tags or tagged resources (in particular in conjunction with the IRW resource ontology designed by Harry Halpin and Valentina Presutti which allows for a fine-grained specification of the different strata of resources encountered on the Web). Then, the link between a tagged resource and a sign used to tag is represented using named graphs. As the declaration of named graphs is not natively supported in RDF, we have integrated the

⁵ <http://vocamp.org/wiki/VoCampNiceSeptember2009>

model from Carroll et al. [3] and the RDF/XML Source declaration syntax from [5]. This choice allows us to express taggings with the different current models of tags (such as SCOT⁶, CommonTag⁷, etc.) and, thus, to bridge them in a very efficient way, and to query several conceptualisations of tags at a time. In addition any description of the act of tagging itself (provenance, means or other contextual and pragmatic information) can be attached to the named graphs of the tag action.

The paper is organized as follows. In section two we discuss the motivations for proposing a new model of tags which tries to account for the wealth of expression hidden behind the simplicity of tags. In section three we detail our modelling of tags and our implementation of their declaration as named graphs, and give some examples before concluding in section four.

2 Nature and usages of tags

For quite a long time, the nature of tags has been partly obfuscated. Thanks to the work accomplished in the wake of the identity crisis faced by the semantic Web (due to the shift from a Web of documents towards a Web of things, URIs being used to identify both "things" and Web pages thus fostering ambiguity) by Patrick Hayes and Harry Halpin on Web proper names formerly known as URIs (now IRIs), *reference* was eventually formally distinguished from *access*⁸ - the name Patrick Hayes has given to this other relationship.

As much *material* devices as *semiotic* ones, tags exhibit a similar kind of duality. To conflate the two aspects would therefore be tantamount to overlooking the simple fact that symbolic bonds between words and things, for instance, do not the least require to be technically implemented in any way. No technical apparatus is necessary for a word to point to an object, no artifact will ever make up for this possibility; in other words, reference pertains to the domain of semantics (or pragmatics, thereof) - not to network engineering.

On the other hand, every tagging system is implemented according to the rules dictated by the needs of the website it is a part of (who's allowed to tag? what? how? etc.). Reference is thus *de facto* complemented on another level by the association, on the technical side of things (the Internet being a physical network where information is exchanged, this should not come out as a surprise) of a *tag* with a *resource* (or rather, if we consider delicious-like tagging, "the representation of a WebResource", as defined by Halpin and Presutti [8]).

⁶ <http://scot-project.org/scot/>

⁷ www.commonitag.org

⁸ Hayes and Halpin [9] convincingly underline the necessity to carefully dissociate the two dimensions in their discussion of URIs. Our main assumption is that it is essential to reckon what this discussion revealed and transpose its result to our endeavour to characterize tags. In a twofold fashion; first, as words or rather potentially meaningful strings of characters; then as a "material" reality granting access to a resource and tightly constrained through limitations attributable to the computerized systems it belongs to (be it the Web or a local application)

Its ins and outs concern interfaces design and, as we previously remarked, the tangible realities of networks and protocols (especially the Web architecture centred around the HTTP protocol and IRIs).

Hence, while the label of the tag itself is nothing more than the string of characters inscribed on it, the latter is also to be construed as a material support belonging to an informational network. While contriving access to a resource, this prop also allows users to add any required bit of texts to whatever resource. It then becomes possible to index, evaluate, share or find again objects that previously overstepped the customary limits of annotation.

The most striking contrast with previous existing bodies of practices and norms, from the point of view of professional indexers, must have been the shift from *a priori*, controlled indexing to these freer forms of annotations. An immediate consequence of this newly acquired freedom, in a nutshell, is that labels are no longer terms of a thesaurus, subject headings in library classifications or even words but all this at the same time and even much more than that. Including, among other things: triple and/or machine tags, URLs, smileys (iconic representations in general), images, messages written in special fonts like Windings, computer code, etc.

If the freedom to choose its own labels that is now offered to users was to be properly acknowledged, then there would remain only one conclusion to be drawn. Contrary to subject headings or descriptors whose semantics is rigidly established in line with a single model of interpretation or through a well-ordered lexicon – thanks to a small number of relations established and postulated in order to evacuate every remainder of ambiguity – tags, understood as inscribed labels, are able to comprise various entities, linguistic or not, thus forbidding all global theorizing on the semantics behind their use. In other words, *the label of a tag is a blank space that is fit to accommodate any sort of inscribable entities. Any sign.* Accordingly, it is therefore completely devoid of any fixed (denotational) semantics (See [12] for a thorough vindication of this point of view. The NEPOMUK ontology NAO⁹, back in 2007, also made a similar point).

3 Modelling tags with the NiceTag ontology

3.1 Tag actions as named graphs

Carroll et al. [3] noted that RDF does not provide mechanisms (apart from statement reification) for talking about graphs and relations between graphs. They introduced Named Graphs in RDF to allow publishers to communicate assertional intent and to sign their assertions. The fact that it is often useful to embody social acts with some record clearly resonates with the scenarios of social tagging. Several authors before them proposed to transform RDF triples into quads [1,4,10,11] appending to them an additional URIref or blank node or ID. The definition of [3] is deliberately simpler than [7] and [15] : “A Named Graph is an RDF graph which is assigned a name in the form of a URIref. The

⁹ <http://www.semanticdesktop.org/ontologies/nao>

name of a graph may occur either in the graph itself, in other graphs, or not at all. Graphs may share URIs but not blank nodes.” [3].

Extending the class `rdfg:Graph` defined in Carroll et al. [3], we define a subclass of named graphs called `TagAction` class and embodying the acts of tagging (see fig. 1). The triples contained in the named graph describe the link between a tagged resource and a sign, as described in fig 1 where two `rdfs:Resource` are linked with the property `nicetag:hasSign`. From this point, our model can support different ways of modelling tagged resources and signs used to tag.

Regarding the model of tagged resources, Halpin & Presutti [8] addressed the fuzziness typical of the notion of a “Resource” as defined by Berners-Lee et al.[2] and its relation to URIs, and proposed a model for solving the identity crisis of resources on the Web. Their model is particularly useful to distinguish between taggings of non-information resources (as when tagging the Eiffel Tower itself, i.e. the physical object, even when doing so through a web page) from taggings of information resources (as when tagging a web page about the Eiffel tower). Thus, if one wants to tag the web page `www.tour-eiffel.fr` to comment on its usefulness for planning a trip, the tagged resource could be modelled with the class `irw:WebResource` from the IRW ontology¹⁰. The signs used to tag can be modelled with all the other currently available models of tags such as SCOT, NAO, Newman’s Tag Ontology¹¹, or CommonTag.

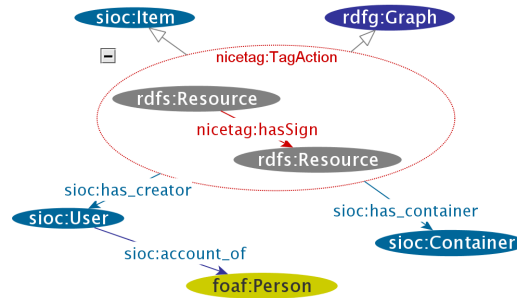


Fig. 1. TagAction instances are declared as named graphs

To account for the nature of the different possible tag actions, we defined subclasses of the `TagAction` class. These subclasses may be of interest to distinguish between tagging performed by machines (`AutoTagAction`), from tagging performed by humans (`ManualTagAction`), from more complex types of tagging as those involving machine tags (`MachineTagAction`). Tag actions may also be collective (`CollectiveTagAction`) or individual (`IndividualTagAction`). Finally, the `TagAction` class is declared as a subclass of the class `Item` from the

¹⁰ <http://ontologydesignpatterns.org/ont/web/irw.owl>

¹¹ <http://www.holygoat.co.uk/owl/redwood/0.1/tags/>

SIOC¹² ontology in order to account for the shareable nature of tags, which can be seen as some sort of post in an online community platform. This, in turn, allows us to describe the place where tag actions are stored with the sioc class `sioc:has_container`, and also the account (`sioc:User`) of the user (`foaf:Person`) of the tag with `sioc:has_creator`. Of course the idea is also to extend that model as new types of tagging acts are identified.

3.2 Modelling tag usages

Current models of tagging aim at linking tags to well defined meanings in order to face the problem of polysemy of tags[13]. Still, polysemy is not the only ambiguity of tags: some meaning resides in the (so far implicit) kind of relationship between the resource and the sign. For example, the use of tag “blog” can assume at least two different meanings, with respect to the same definition of the word “blog” it can mean that a resource *is about* blogs, or that a resource *is a* blog. Moreover, some tags are intended for *personal* use and to only make sense for the applier.

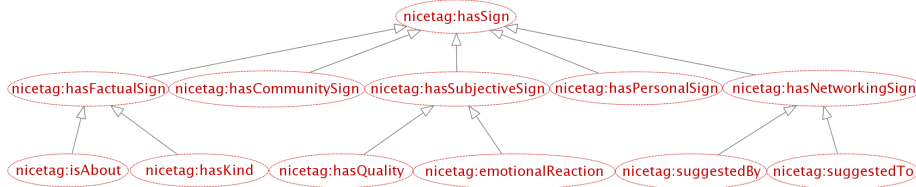


Fig. 2. `nicetag:hasSign` sub-properties

Inspired by previous studies, and in particular by Golder & Huberman [6], we modelled the different possible uses of tags with sub-properties of the `nicetag:hasSign` property; we also grouped them in three broader classes: *factual*, *subjective* and *personal*, as proposed in [14]. Beyond this classification, we added two more branches to represent community and networking tags (see fig. 2).

The property `isAbout` represents the most common use of a tag, to identify the *topic* of an item, whereas `hasKind` is intended for all cases in which a tag is used to define *what a resource is* (e.g.: “forum”, “video”); both are subsumed by `hasFactualSign`, the first subproperty of `hasSign`. The second branch, for subjective tags, comprises two subproperties: `hasQuality`, to associate a resource with an adjective or with any kind of sign expressing a quality (e.g.: “nice”, “bullshit”), and `emotionalReaction`, for tags expressing an emotion stirred up by a resource; typical examples are exclamations and smileys (e.g.: “wow!”, “^_^”). The third direct subproperty of `hasSign` is `hasPersonalSign`, which covers all uses of tags intended to just make sense for the applier; this includes Golder & Huberman’s classes *task organizing* (like “toread”) and *self reference* (like

¹² <http://sioc-project.org/ontology>

“mystuff”). Similarly, we introduced the property `hasCommunitySign` for tags that have an intended audience of a community. For example, we used the tag “#vocampnice2009” to share resources about the VoCamp across multiple social Web applications.

As a last branch we’ve added the two properties `suggestedTo` and `suggestedBy`, as subproperties of `hasNetworkingSign`, to model networking tasks. Some bookmarking systems already have a special syntax for this (e.g.: delicious “for:-username” tags).

3.3 Using RDF/XML Source declaration to implement and use named graphs

In SPARQL, when querying a collection of graphs, the `GRAPH` keyword is used to match patterns against named graphs. However the RDF data model focuses on expressing triples with a subject, predicate and object and neither it nor its RDF/XML syntax provide a mechanism to specify the source of each triple. To serialize named graphs, Carroll et al. used TriX and TriG [3] but noted that RDF/XML is the deployed base. Therefore, we proposed in the W3C Member Submission “RDF/XML Source Declaration” [5] an XML syntax to associate to the triples encoded in RDF/XML an IRI specifying their origin ; it uses a single attribute to specify for these triples represented in RDF/XML the source they should be attached to. The IRI of the source of a triple is:

1. the source IRI specified by a `cos:graph` attribute on the XML element encoding this triple, if one exists, otherwise
2. the source IRI of the element’s parent element (obtained following recursively the same rules), otherwise
3. the base IRI of the document.

The scope of a source declaration extends from the beginning of the start-element in which it appears to the end of the corresponding end-element, excluding the scope of any inner source declarations. Such a source declaration applies to all elements and attributes within its scope. If no source is specified, the URL of the RDF/XML document is used as a default source. Only one source can be declared as attribute of a single element.

The example in listing 1.1 shows how this applies to declare a tag as a named graph. Lines 4-7 declare the tag as a graph named `http://mysocialsi.te/tag#7182904` . Lines 8-11 reuses the name of the graph to qualify the tag as a tag created manually by “Fabien Gandon” the 7th of October 2009.

Loading this RDF in a compliant triple store one can then run SPARQL queries like the one in listing 1.2. Line 2 searches for named graphs and the triples they contain. Line 3 enforces these graphs to be manually generated tags.

Thanks to the flexibility of our model, we are able to express tags in many different possible flavors. A sign used to tag can be a mere character string, or an instance of the tag class from SCOT or CommonTag. In the the latter case, the meaning of the tag (given by `ctag:means` property) will be included within

Listing 1.1. Declaration of a tag as a named graph using RDF/XML

```
1 <rdf:RDF xmlns:dc='http://purl.org/dc/elements/1.1/'
2   xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
3   xmlns:cos='http://www.inria.fr/acacia/corese#'>
4   <rdf:Resource rdf:about='http://www.yesand.com/'
5     cos:graph='http://mysocialsi.te/tag#7182904'>
6     <nicetag:isAbout>improvisation</nicetag:isAbout>
7   </rdf:Resource>
8   <nicetag:ManualTagAction rdf:about='http://mysocialsi.te/tag#7182904'>
9     <dc:creator>Fabien Gandon</dc:creator>
10    <dc:date>2009-10-07T19:20:30.45+01:00</dc:date>
11  </nicetag:ManualTag>
12 </rdf:RDF>
```

Listing 1.2. SPARQL query to retrieve tags declared as named graphs

```
1 SELECT ?t ?a ?g WHERE {
2   GRAPH ?tag { ?t ?a ?g }
3   ?tag rdf:type nt:ManualTagAction }
```

the named graph of the tag action. As we mentioned earlier the IRW ontology[8], tagged resources can be modelled with the class `irw:Resource` or its subclasses in order to distinguish properly what is being tagged. As a consequence, it is possible to retrieve in a single query all taggings expressed with our model, regardless the type of signs used to tag or the type of the tagged resource.

4 Conclusion

In this paper we have proposed a general and flexible model which represents tags, taking into account the "speech act" dimension of tagging, by means of named graphs. The essence of a tag in NiceTag is to embody in a record one or more RDF triples associating a resource with a sign and this core information can be enriched in several directions. To allow for the specification of the function of a tag, we have created several subproperties to cover different kinds of relationships between the sign and the tagged resource. The named graph containing the tag is itself an instance of the class `TagAction`, and can thus have properties associated with it (like the tagger, the tagging date, etc.). Moreover, it is possible to define the kind of a `TagAction` by choosing one of the subclasses we have defined. All these primitives (sign classes, function properties, tag action classes) are also designed to be extended at will.

This way, thanks to the use of the RDF/XML Source Declaration syntax to assign a URI to a tag action, we obtain a full expressive richness to represent tags from a multiplicity of facets, avoiding the burden of RDF reification. Both the Named Graphs model and the RDF/XML syntax extension provide a high-value for a small, incremental and backward-compatible change to the Semantic Web Recommendations. Combined with the tagging vocabularies this model

provides us with a very flexible and extensible framework for social tagging interoperability.

Existing tagging ontologies can be integrated into this model on the side of the “sign”, allowing, for instance, for the association of a tag with a well defined concept with MOAT¹³ or CommonTag, or for the specification of semantic relationships between tags with SCOT. On the side of the resource being tagged, we have pointed out the problems raised by the manifold nature of URIs in the semantic Web, where it can be often unclear if a tag should be referred to a Web resource or to the real object or entity it represents, and we have suggested the use of the IRW ontology in combination with NiceTag to face the *identity crisis*.

References

1. D. Beckett. Redland notes - contexts. <http://www.redland.opensource.ac.uk/notes/context.html>, 2003.
2. T. Berners-Lee, R. Fielding, and L. Masinter. Uniform resource identifiers (uri): Generic syntax. <http://www.ietf.org/rfc/rfc3986.txt>, 2005.
3. Jeremy J. Carroll, Christian Bizer, Pat Hayes, and Patrick Stickler. Named graphs, provenance and trust. In *14th Int. Conference on World Wide Web WWW*, pages 613–622, New York, NY, USA, 2005. ACM.
4. E. Dumbill. Tracking provenance of rdf data. Technical report, ISO/IEC, 2003.
5. Fabien Gandon, Virginie Bottolier, Olivier Corby, and Priscille Durville. Rdf/xml source declaration, w3c member submission. <http://www.w3.org/Submission/rdfsouce/>, 09 2007.
6. Scott A. Golder and Bernardo A. Huberman. Usage patterns of collaborative tagging systems. *J. Inf. Sci.*, 32(2):198–208, 2006.
7. R. M. R. Guha and R. Fikes. Contexts for the semantic web. In *Int. Sem Web Conf, ISWC*, 2004.
8. Harry Halpin and Valentina Presutti. An ontology of resources: Solving the identity crisis. 5554:521–534, 2009.
9. Patrick J. Hayes and Harry Halpin. In defense of ambiguity. *Int. J. Semantic Web Inf. Syst.*, 4(2):1–18, 2008.
10. Intellidimension. Rdf gateway - database fundamentals. <http://www.intellidimension.com/pages/dfgateway/dev-guide/db/db.rsp>, 2003.
11. R. MacGregor and I.-Y. Ko. Representing contextualized data using semantic web tools. In *Practical and Scalable Semantic Systems (ISWC workshop)*, 2003.
12. Alexandre Monnin. Tags and folksonomies as artifacts of meaning. *Philosophy of Engineering and Artifact in the Digital Age. A First Eastern European (Romanian) Perspective*, Cambridge Scholar Publishing, 2009.
13. Alexandre Passant and Philippe Laublet. Meaning of a tag: A collaborative approach to bridge the gap between tagging and linked data. In *WWW Workshop on Linked Data on the Web (LDOW)*, Beijing, China, Apr 2008.
14. Shilad Sen, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper, and John Riedl. Tagging, communities, vocabulary, evolution. In *20th conf. Computer Supported Cooperative Work, CSCW*, pages 181–190, New York, NY, USA, November 2006. ACM.
15. M. Sintek and S. Decker. Triple - a query, inference, and transformation language for the semantic web. In *Intl. Sem. Web Conf., ISWC*, 2002.

¹³ <http://moat-project.org/>